

基于 DRFP 网络的无人机对地车辆目标识别算法

张钟毓^{1,2,3,4,5}, 刘云鹏^{1,2,4,5}, 王思奎^{1,2,3,4,5}, 刘天赐^{1,2,3,4,5}, 林智远^{1,2,3,4,5}

- (1. 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;
2. 中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110016;
3. 中国科学院大学, 北京 100049;
4. 中国科学院光电信息处理重点实验室, 辽宁 沈阳 110016;
5. 辽宁省图像理解与视觉计算重点实验室, 辽宁 沈阳 110016)

摘要: 针对无人机在复杂战场环境的侦察任务中, 目标在视场中尺寸过小、边缘和纹理信息较少所造成的目标识别难题, 提出一种新的基于深度学习的单阶段目标识别网络 DRFP。DRFP 网络以残差结构为骨架, 使用特征金字塔结构实现特征融合; 其次在损失函数中使用添加了调整因子的交叉熵函数, 实现对难样本的重点关注、训练; 最后使用高斯型非极大值抑制算法(G-NMS), 提高目标密集区检出率。使用无人机航拍图像数据集进行地面车辆目标识别的实验结果表明: 所提出的单阶段模型的精度(mAP 值)为 83.16%, 达到了两阶段网络模型的水平; 同时, 识别速度符合实时性的要求。

关键词: 小目标识别; 无人机图像; 深度学习

中图分类号: TP391 **文献标志码:** A **DOI:** 10.3788/IRLA201948.S226001

Vehicle target recognition algorithm for UAV image based on DRFP

Zhang Zhongyu^{1,2,3,4,5}, Liu Yunpeng^{1,2,4,5}, Wang Sikui^{1,2,3,4,5}, Liu Tianci^{1,2,3,4,5}, Lin Zhiyuan^{1,2,3,4,5}

- (1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;
2. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110016, China;
3. University of Chinese Academy of Sciences, Beijing 100049, China;
4. Key Laboratory of Opto-electronic Information Processing, Chinese Academy Sciences, Shenyang 110016, China;
5. The Key Lab of Image Understanding and Computer Vision, Shenyang 110016, China)

Abstract: In order to solve the problem of small target recognition caused by small size, less edge and texture information in the field of view for UAV in complex battlefield environment, a new model based on deep learning for small target recognition Deep Residual and Feature Pyramid (DRFP) was proposed in this paper. Firstly, the residual structure was used as the skeleton of the model, and the feature pyramid structure was used to achieve feature fusion. Secondly, the cross-entropy function with adjusting factor was used in the loss function to realize the focus of attention on difficult samples. Finally, a non-maximum Gaussian suppression algorithm was used to improve the detection rate of target-intensive areas.

收稿日期: 2019-04-01; 修订日期: 2019-05-14

基金项目: 中国科学院国防科技创新基金(Y8K4160401)

作者简介: 张钟毓(1996-), 男, 硕士生, 主要从事目标检测方面的研究。Email: zhangzhongyu@sia.cn

导师简介: 刘云鹏(1980-), 男, 研究员, 硕士生导师, 博士, 主要从事图像处理与模式识别方面的研究。Email: ypliu@sia.cn

The experimental results show that the accuracy (mAP) of proposed single stage model is 83.16% using UAV-images towards vehicle recognition, which achieves the level of two stage network model. At the same time, the recognition speed meets real-time requirements.

Key words: small target recognition; UAV-image; deep learning

0 引言

现代战争中,无人机通过搭载传感器实现对战场动态信息的实时监测,地面目标识别是其中极为重要的一环。无人机的拍摄视角使得地面目标尺寸较小、边缘和纹理信息较少,给目标识别任务带来了极大的挑战。传统的模板匹配方法在识别准确度和速度上无法匹配现代战争的要求,探索新方法解决该问题变得十分迫切。

随着大数据、机器学习、人工智能技术的迅猛发展,以大规模训练样本库为基础的深度学习方法取得了令人瞩目的成绩。在计算机视觉领域,基于卷积神经网络(CNN)的各种深度学习网络框架在图像质量评价^[1]、机器人技术^[2]及其他生产生活领域^[3]中取得了巨大的成功,识别准确度逐步追上并超越人眼水平。现有的基于深度学习的目标识别方法主要分为两阶段方法和单阶段方法。两阶段方法以 R-CNN^[4-5]系列网络为代表,第一阶段的任务是在图片中选出感兴趣的区域并按照正负样本加以区分;第二阶段对选出的正样本区域进行分类、定位的多任务损失计算,从而得到检测结果。单阶段方法以 Yann LeCun 等人提出的 OverFeat 算法为先驱,利用滑动窗口、anchor box 等方法直接获得候选区域,对候选区域进行回归操作得到其定位结果。相比较而言,两阶段方法能由第一阶段筛选出质量更高的前景区域,所以识别准确度更高但是速度较慢;单阶段方法的识别速度极快,但是定位结果不如两阶段方法准确。YOLOv3^[6]作为当下性能优异的单阶段目标识别网络,不仅在识别速度上达到实时性要求,而且在识别准确度上与两阶段网络相比也相差不多。综合考虑上述问题,笔者决定以 YOLOv3 网络结构为主要借鉴,设计一种新的基于深度学习的单阶段网络模型 DRFP。面向常见场景所使用的车辆数据集中的图像有着距离目标较近、目标在视场中比例较大、

边缘和纹理清晰可见等特点;而无人机获取的图像中,地面目标更小、环境更为复杂、边缘和纹理信息更少。因此,在无人机目标识别领域取得突破的关键点在于小目标识别能力的加强。同时,无人机视角造成了较多的识别得分较低的正样本与得分较高的负样本(难样本)以及较密集的目标区域。为了解决上述问题,文中算法(1)在损失函数中使用添加了调整因子的交叉熵函数,实现对难样本的重点关注,提高识别准确度;(2)以定位置信度为导向,使用高斯型非极大值抑制算法提高密集区域的识别准确度。

文中实验所使用的数据集是航空影像中的车辆检测(Vehicle Detection in Aerial Imagery,VEDAI)数据集,VEDAI 作为小目标数据集的代表之一受到了广泛关注,大量模型使用该数据集进行实验验证。其中 Terrail J O D^[7]等人构建的 Faster RER-CNN 网络是识别效果最好的模型之一,Faster RER-CNN 是在 Faster R-CNN 的基础上针对小目标识别任务进行的改进,在实验部分笔者同时使用 Faster R-CNN 和 Faster RER-CNN 作为两阶段模型的代表与 DRFP 进行比较,以验证文中算法在满足实时识别的同时识别精度达到两阶段网络的水平。

1 基于 CNN 的 DRFP 及构建方法

1.1 DRFP 的结构模型

DRFP 的结构设计以更准、更快地识别小目标为出发点,利用残差模块加深网络的优势来提取更加抽象的图像信息;使用特征金字塔融合多个特征层的特点改善小目标识别效果;采用多层稀疏的结构设计来缓解大量使用残差模块带来的参数大幅增多问题。从图 1 可以看出:DRFP 使用了 1 个含有 32 个 3×3 卷积核的卷积层(Conv)、6 个参数设置不同的残差模块(Res1~Res6)、2 个特征融合层(Concat)、1 个使用 Dedicated Loss 产生初步结果的识别层、1 个高斯

型非极大值抑制算法层, 最终结果标记在输出图像 (OUT) 中。这样的“多层”结构设计使得网络具有更强的非线性表达能力, 以提高识别准确性。“稀疏”

意味着相较于其他单阶段网络模型(如 YOLOv3), DRFP 各层中所使用的卷积核数量少, 具体参数设置见图 1。

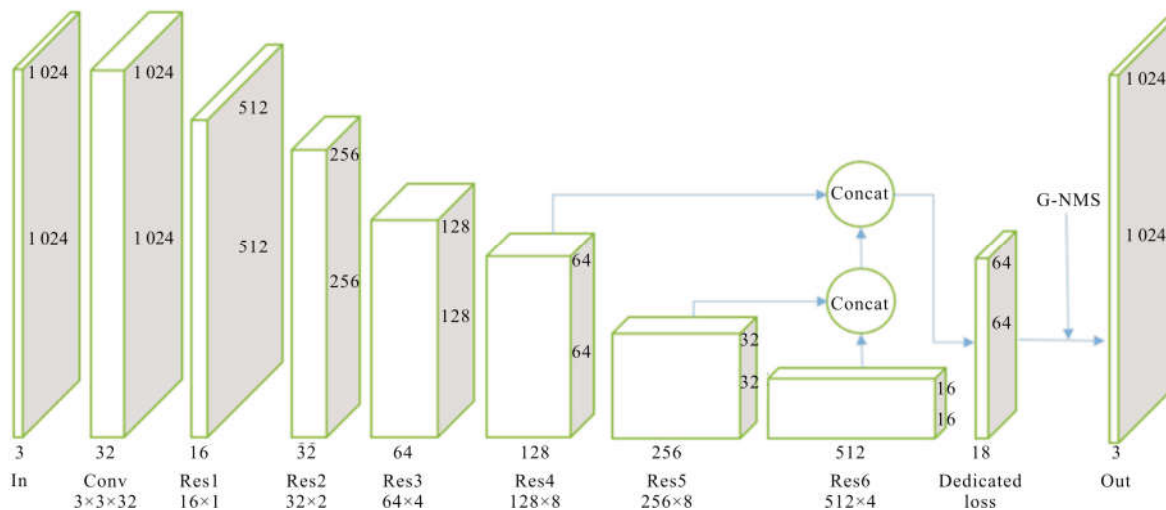


图 1 DRFP 网络结构图

Fig.1 Network structure of DRFP

使用池化层实现下采样会丢失一定的位置信息, 因此, DRFP 采用全卷积结构, 池化效果通过将卷积层的步长设为 2 来实现。在卷积层中使用激活函数 Leaky ReLU:

$$y = \begin{cases} x & \text{if } x > 0 \\ \frac{x}{a} & \text{if } x \leq 0 \end{cases} \quad (1)$$

式中: a 的取值为 10。残差模块 (Res1~Res6) 使用 ReLU 作为激活函数, 卷积核数目从 16 开始, 后一个残差模块卷积核数目翻倍, 直到 512 为止(图 1 中与卷积核数目并列放置的数字代表该残差模块的重复次数)。In 表示输入的图像数据, 初始输入的图像为 1024x1024 分辨率, 含有 RGB 三通道。在实验中, 识别层位于尺寸为 64x64 的特征层后, 每一个特征点设置 9 个尺度不同的初始候选框。识别层所得到的初步识别结果经过 G-NMS 处理, 得到最终的识别结果并展现在输出图像中。模型主要由 6 类(Res1~Res6)、共计 27 个残差模块组成, 残差结构的提出是为了解决越深的网络越难训练的网络退化问题(主要是梯度消失和梯度爆炸问题), 即随着网络加深, 通过反向传播算法计算得出的参数更新梯度值将趋于 0 或一个很大的值。梯度消失和梯度爆炸问题使

得深层网络并不能自主实现多余网络结构的自主恒等映射。为了解决这个问题, 何凯明等人在参考文献[8]中提出了一个类似“短路”的“残差模块”, 给了网络一个至少不会变差的通路, 图 2 中由输入“ x ”到“+运算”的连接线代表着这样的一个通路。使用残差模块设计网络模型的一个重要设计原则是: 当特征图大小降低一半时, 特征图的数量增加一倍以保持网络层的复杂度。

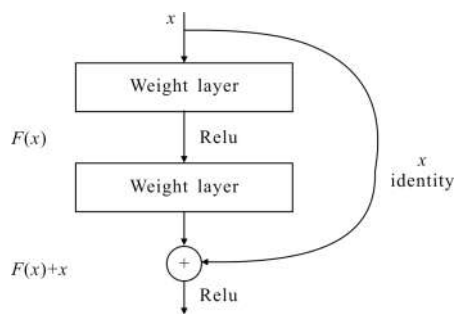


图 2 残差模块结构图

Fig.2 Residual module structure

1.2 特征融合金字塔模型

近年来针对各级特征层的可视化研究让笔者认识到: 感受野小的低层特征由于结构的简单而只能

提取图像的边缘、轮廓、纹理等信息,但是定位信息较为准确;高层网络能实现更加复杂的空间表示,从而实现更加抽象的图像概念判断并以此来识别物体,但是位置信息丢失严重。上述情况导致了低层的特征语义信息比较少,但是目标定位准确;高层的特征语义信息比较丰富,但是目标定位较为粗略。对 VEDAI 所有已标注的目标定位框求得的平均值为 32×20,这就意味着待识别的目标平均占据 32 pixel×20 pixel。在此背景下,感受野普遍大于 32×32 的高层特征层对于小目标的定位是不利的,而感受野小的低层特征中包含的图像信息又不足以支撑目标识别。

为了解决上述问题,DRFP 采用特征金字塔结构实现特征融合,特征金字塔通过简单的网络连接改变,在基本不增加原有模型计算量的情况下,大幅度提升了小物体识别的性能。具体做法是将感受野大的高层特征层进行上采样,与经过 1×1 卷积层的低层特征层进行融合,这里的融合一般是指并列排列(图 1 中 Concat 部分),如 128 层高层特征与 256 层低层特征融合之后得到 384 层特征层。此时得到的特征层既包含有图像的高层语义信息,又包含有图像的低层纹理、轮廓等信息,可以理解为高层语义信息指导低层特征进行分类、定位等。DRFP 只使用合适尺寸(64×64)的特征融合层进行识别操作,这样做的好处是不仅可以减少作用小的识别层带来的计算量,而且避免了他们对于参数更新方向的“误导”。

2 难样本损失函数设计

损失函数对基于深度学习的识别网络来说,不仅是识别阶段中候选目标的判别标准,更在网络训练阶段起着决定性的作用,关乎网络的最终识别性能。DRFP 损失函数的结构使用了当下主流的由物体判别损失、分类损失和定位损失三部分组成的损失函数结构:

$$\sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} (C_i + \hat{C}_i)^2 + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{noobj} (C_i + \hat{C}_i)^2 + \sum_{i=0}^{S^2} 1_{ij}^{obj} \sum_{classes} (p_i + \hat{p}_i)^2 + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [\sqrt{(w_i + \hat{w}_i)^2} + \sqrt{(h_i + \hat{h}_i)^2}] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(x_i + \hat{x}_i)^2 + (y_i + \hat{y}_i)^2] \quad (2)$$

式中:前两项中的 obj 与 noobj 分别用来计算判别为

目标(前景)和非目标(背景)的损失,不包含前景的候选区域样本即为负样本;后两项代表定位损失,包括目标中心点横纵坐标及候选框的宽度、高度信息;第三项为分类损失。类别越多、正样本比例越小识别结果越容易出现偏差,这是识别过程中的难点,也是本模型改进损失函数的主要目的所在。分类损失项中的 $\hat{p}_i(c)$ 普遍使用公式(3)所示的 sigmoid 函数激活后的交叉熵(CE)函数,其中 p 为其类别得分。

$$CE(p, y) = \begin{cases} -\log(p) & \text{if } y=1 \\ -\log(1-p) & \text{others} \end{cases} \quad (3)$$

无人机平台采集到的图像有如下特点:(1) 目标相对较小、纹理信息缺失,导致正样本的分类置信度较低从而成为难样本;(2) 目标在图像中占据的空间小,导致候选框中负样本占大多数,但是其中大多是简单样本,继续使用传统的交叉熵函数则无法去除大量简单样本的影响。DRFP 的损失函数充分考虑了以上难点,将重点训练难样本。Dedicated Loss 中 $\hat{p}_i(c)$ 使用改进后的交叉熵函数:

$$DCE(p, y) = \begin{cases} -\log(p)(1-p)^2 & \text{if } y=1 \\ -\log(1-p)p^2 & \text{others} \end{cases} \quad (4)$$

对改进前后的函数分别进行求导,并代入 $y=1$ (分类正确)的情况计算其反向传播时的参数更新值,计算结果如表 1 所示。可以看出,改进后的损失函数通过加大难样本($y=1$ 时, P 较小)对于参数更新的影响,来加大对于难样本的训练力度,对于简单样本($y=1$ 时, P 较大)则关注度降低。

表 1 两种反向传播算法的更新值对比

Tab.1 Update values comparison of two kinds of backpropagation algorithm

$P(\text{positive})$	Cross entropy	Dedicated loss
0.1	0.900	1.102
0.2	0.800	0.924
0.3	0.700	0.696
0.4	0.600	0.479
0.5	0.500	0.298
0.6	0.400	0.162
0.7	0.300	0.071
0.8	0.200	0.022
0.9	0.100	0.002

3 高斯型非极大值抑制算法

在目标识别算法的结束阶段,非极大值抑制(NMS)是一个至关重要的步骤,然而传统的非极大值抑制算法有着容易造成定位信度低、密集区域漏检等问题。下面将介绍 DRFP 中的高斯型非极大值抑制算法,从而改善上述两个问题。

3.1 提高定位信度优先级

目标识别准确性由分类准确性和定位准确性两部分组成,传统 NMS 算法使用分类置信度作为候选框的排序标准,在损失函数使用 Dedicated Loss 实现对分类准确性提高的情况下,继续使用传统 NMS 算法会造成模型太过看重分类置信度而忽视了定位准确性。因此在设计 G-NMS 时,DRFP 更加注重提高定位准确性的优先级,将评价候选框的首要指标从分类置信度变成定位信度,这里的定位信度的意义是候选框

与目标真实定位框之间的交并比(IOU)值在输入图像进行目标识别的过程中,不含有目标真实定位框的坐标信息,因此,要求网络另外训练出预测候选框与真实定位框之间 IOU 值的能力。为了实现网络预测 IOU 值的能力,使用 IOU-Guided 的网络结构,DRFP 将公式(2)中的 \hat{C}_i 进行拆分。 \hat{C}_i 原本表示为:

$$\hat{C}_i = \text{Pr}(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} \quad (5)$$

也就是候选框中含有物体的概率与候选框与 Ground Truth 之间 IOU 值的乘积。在训练过程中,只提取 \hat{C}_i 中的 IOU 部分作为参数调整依据,训练目标是使 IOU 值趋向于 1。在此基础上,G-NMS 中使用预测的 IOU 作为选择非极大值的指标。

3.2 解决密集区域漏检问题

非极大值抑制可以很大程度上避免同一目标被重复识别情况的发生,其核心思想是 IOU 比较大的候选框视为对同一个目标的识别结果,在这些候选框中只保留拥有最大置信度的那一个。被采用较多的传统非极大值抑制算法的流程为:

步骤 1 取出预测为某一个类别的所有候选框,按照该类别预测的置信度对每个候选框排序。

步骤 2 比较当前拥有最大置信度的候选框 M

与其余所有候选框的相交部分与并集的比,IOU 值超过阈值 N 的按照公式(6)计算:

$$s = \begin{cases} s & \text{iou}(M, b) < N \\ 0 & \text{iou}(M, b) \geq N \end{cases} \quad (6)$$

被抑制,通过公式(6)可以看出,传统方法中的抑制代表被直接删除。

步骤 3 将 M 取出并找出新的 M ,剩余的候选框重复进行步骤 2 的操作。

每一个类别都会进行一次步骤 1~3 的非极大值抑制操作。

在实验过程中发现,将与 M 之间 IOU 值大于等于阈值的候选框直接删去会产生密集区域的漏检情况。为了解释产生这种现象的原因,假设两个地面车辆 c_1 、 c_2 的定位信度分别是 0.95 和 0.9,同时两辆车由于相距很近或 c_1 较长且倾斜停放,那么此时 c_1 的候选框就很容易与 c_2 的候选框产生大于阈值的 IOU 值。依据传统非极大值抑制算法,定位信度为 0.9 的 c_2 会被直接删去,从而形成漏检。为了避免这种情况的发生,将公式(6)替代为高斯型公式:

$$s = s * e^{-\frac{\text{iou}(M, b)^2}{0.6}} \quad (7)$$

这样优化的好处是并不是直接将非极大值候选框删去,而是降低他们的置信度,并给与继续参与非极大值抑制的机会。若是在 G-NMS 流程结束之后这些候选框的置信度高于设置的识别阈值,那么它们也会被作为识别结果进行输出。

将 G-NMS 算法流程最终确定为:

步骤 1 首先取出预测为某一个类别的所有候选框,按照定位信度(预测的 IOU 值)给每个候选框排序。

步骤 2 按照公式(7)计算所有候选框新的置信度得分。

步骤 3 之后将 M 取出并找出新的 M ,剩余的候选框重复进行上一步操作。

每一个类别都会进行一次非极大值抑制操作,得到最终的识别结果。

4 实验结果与分析

4.1 实验设置

实验使用的是航空影像中的车辆检测(VEDAI)数据集,VEDAI 数据集含有 1 251 张无人机采集到

的地面车辆图像,每一张图像分辨率均为1 024×1 024。所有的图像均采取了与地面相同的距离,地面采样距离(GSD)为 12.5 cm/pixel(cmpp)。VEDAI 数据集对地面车辆进行了详细分类,分成了小轿车、卡车、野营车、公交车、皮卡、拖拉机等类别,一些诸如推土机的少量样本被标为其他。不仅是车辆,停在陆地上的小船和大型目标如飞机也被标注出来。实验中笔者沿用该分类。在针对图像以及标注文件质量的检查中,删去 4 张信息不全的图像及其标注文件,最终使用的样本集容量为 1 247。将这些图像分为 847 个样本组成的训练集、300 个样本组成的验证集以及 100 个样本组成的测试集。在训练过程中使用 NVIDIA 旗下的 GeForce GTX 1080Ti,学习率设为 0.001,每一次迭代训练使用的图片数量设为 32。

实验分为三部分,实验一、实验二分别验证 Dedicated Loss、G-NMS 对识别性能的提高程度。实验三验证融合了 Dedicated Loss 和 G-NMS 的单阶段模型 DRFP 在识别准确度和速度上能否达到以下两点:(1) 目标识别准确度达到现有的两阶段网络的水平;(2) 目标识别速度保持单阶段网络实时性的优势。

实验三的对标对象为单阶段网络的代表 YOLOv3 以及两阶段网络的代表 Faster R-CNN,在此基础上引入 Faster RER-CNN。Faster RER-CNN 是以 Faster R-CNN 为基础针对小目标进行改进的成果,在小目标识别准确度方面具有优势。在实验一、二中,评价标准是召回率,召回率表示正样本被识别出的概率。为了更加全面地进行比较,实验一、二都进行了 5 组实验,将识别结果与真实值之间的 IOU 阈值依次设置为 0.5、0.6、0.7、0.8、0.9。实验三

的评价指标为各类平均精确度的均值 (mAP)、网络层数、参数数量和识别速度。

4.2 实验结果

4.2.1 Dedicated Loss 与传统损失函数对比

实验一是 Dedicated Loss 与交叉熵(CE)函数的对比实验,二者的召回率对比如表 2 所示。观察表 2 可以看出:Dedicated Loss 在实验一的 5 组对比实验中表现均优于传统损失函数,尤其在 IOU 阈值为 0.6 时,召回率较传统损失函数提高了 3.3 个百分点。这说明在 Dedicated Loss 的帮助下,DRFP 对难样本进行了有针对性的训练,参数更新方向更有利于难样本的识别。

表 2 两种损失函数的召回率对比

Tab.2 Recall comparison of two loss functions

Recall	IOU				
	0.5	0.6	0.7	0.8	0.9
CE	85.7%	77.2%	61.2%	38.3%	9.8%
Dedicated Loss	88.2%	80.5%	63.1%	40.1%	10.5%

具体的识别结果如图 3 所示,第二行图片是对应识别结果的局部放大,以便于观察。通过图 3 可以看出:在城市(图 3(a))、荒漠(图 3(b))、农田(图 3(c))等多种背景下,使用 Dedicated Loss 的模型均有着出色的目标识别能力。其中,城市背景下较难处理的树木、房屋遮挡(图 3(d))问题,以及荒漠背景下车辆与荒漠颜色相近(图 3(e))等问题都得到了较好解决。识别结果图表明:Dedicated Loss 并没有以牺牲准确率为前提提升召回率,识别结果中的误检情况几乎没有发生。





图 3 实验一的部分识别结果

Fig.3 Part of recognition result in experiment 1

4.2.2 G-NMS 与传统非极大值抑制算法对比

实验二是 G-NMS 和传统非极大值抑制算法的效果对比, 两种非极大值抑制算法的召回率对比如表 3 所示。观察表 3 可以看出: G-NMS 在一定程度上避免了目标密集情况下的漏检, 由于 VEDAI 中密集目标的样本不多, 故而实验结果所示的性能提升有限。对比发现: G-NMS 对于减少目标密集区域的漏检有着一定帮助。

表 3 两种非极大值抑制算法的召回率对比

Tab.3 Recall comparison of two NMS

Recall	IOU				
	0.5	0.6	0.7	0.8	0.9
NMS	85.7%	77.2%	61.2%	38.3%	9.8%
G-NMS	87.2%	78.5%	61.9%	39.1%	10.2%

4.2.3 DRFP 与具有代表性的网络对比

实验三是 DRFP 与其他具有代表性的网络之间的性能对比, 对比对象为两阶段模型的代表 Faster R-CNN、Faster RER-CNN, 以及单阶段模型的代表 YOLOv3。各种网络性能的量化分析结果如表 4 所示。通过量化分析表可以看出来: 两阶段模型(Faster R-CNN、Faster RER-CNN)虽然在识别准确度(mAP)方面具有一定的优势, 但是由于其网络结构复杂、参数量较多, 造成识别速度相比于单阶段网络模型较慢; 而单阶段模型如 YOLOv3 具有识别速度快、网络结构简单、参数数量少等特点, 但是其识别准确度与面向小目标的两阶段模型相差较大; 对比上述两种

模型, 文中提出的 DRFP 网络模型, 多层稀疏的结构设计使得其参数数量大幅减少、识别速度大幅提高, 同时, 损失函数和非极大值抑制算法的改进, 使其具有与两阶段网络模型相差不多的识别准确度。

表 4 各种网络性能的量化分析

Tab.4 Quantitative analysis of the performance of various networks

Models	mAP	Parameters/million	Speed/FPS
Faster R-CNN	80.69%	138.5	6.33
Faster RER-CNN	84.18%	More than Faster R-CNN	2.74
YOLOv3	79.71%	61.9	33
DRFP	83.16%	19.2	57

为了更为直观地展现 DRFP 的良好性能, 图 4 给出了 DRFP 在无人机拍摄到的图像上的部分识别结果, 其中第一列图片是使用 DRFP 的识别结果, 第二、第三列图片是对识别结果的局部放大, 以便于观察。观察图 4 可以看出: 使用文中提出的 DRFP 网络进行无人机图像目标识别, 其识别结果是十分令人满意的。在将地面车辆目标较好识别出的情况下, 错检情况极少, 说明该算法无论是召回率还是准确率都达到了较高的水平。无论是较小的汽车目标, 还是相对更大的货车、房车、飞机等地面目标, DRFP 都能得到较好的识别结果。从中可以看出: DRFP 具有较强尺度自适应能力, 能够提取多尺度的目标特征信息, 从而识别出尺度大小跨度较大的地面目标。

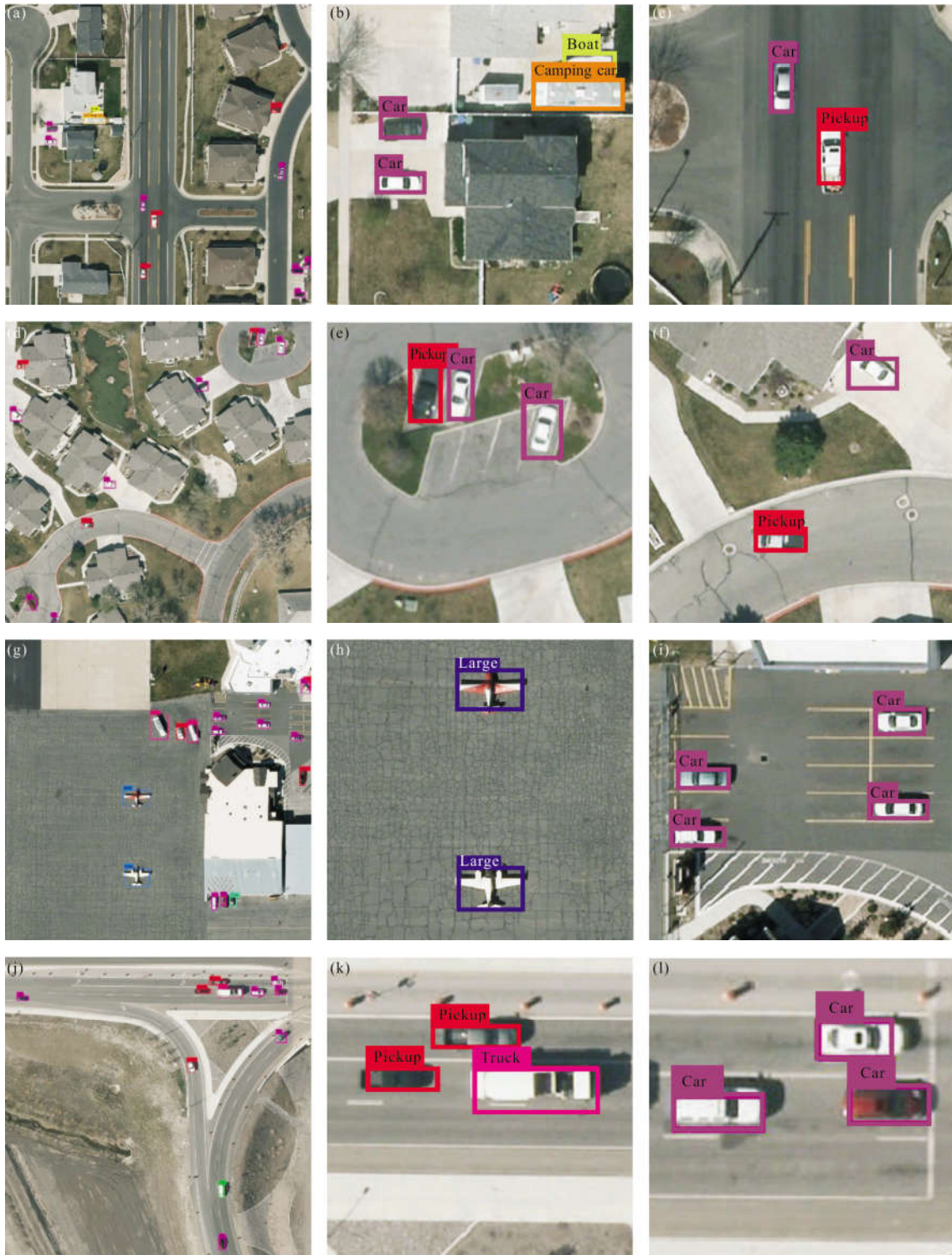


图 4 DRFP 在无人机图像上的部分识别结果

Fig.4 Part of recognition results of DRFP on UAV-image

5 结论

文中提出了一种基于 DRFP 网络的无人机对地车辆目标识别算法,针对无人机图像中目标较小、边

缘和纹理特征较少的特点,设计了以残差模块为主要架构、使用特征金字塔融合特征的多层稀疏网络模型,在此基础上使用 Dedicated Loss、G-NMS 筛选出最优的识别结果,最终实现了面向无人机拍摄视

角下的地面车辆目标的准确、快速识别。在 VEDAI 数据集中的实验结果表明:文中所提出算法不仅在识别准确度方面达到了两阶段网络模型的水平,识别速度也达到了实时识别的程度。同时,DRFP 算法的识别速度相比于以识别速度见长的单阶段模型 YOLOv3,得到了大幅提升。

将深度学习应用到无人机采集到的图像中是一项很有趣也很有挑战的研究课题,对于加强中国人民解放军的战场侦查、打击能力有着深远影响。未来的工作将致力于在文中提出的 DRFP 网络的基础上,进一步提升对于无人机视角下的地面目标尤其是军事目标的识别精度和速度,特别是在模型运行平台方面,计划借鉴卷积核改进^[9-10]、知识蒸馏等模型加速思路,实现在结构较简单的嵌入式开发板的计算力限制下的实时、准确识别。

参考文献:

- [1] Yao Wang, Liu Yunpeng, Zhu Changbo. Deep learning of full-reference image quality assessment based on human visual properties [J]. *Infrared and Laser Engineering*, 2018, 47(7): 0703004. (in Chinese)
- [2] Yu Yi, Kong Lingbao, Zhang Haitao, et al. An improved material removal model for robot polishing based on neural networks [J]. *Infrared and Laser Engineering*, 2019, 48(3): 0317005. (in Chinese)
- [3] Geng Lei, Liang Xiaoyu, Xiao Zhitao, et al. Real-time driver fatigue detection based on morphology infrared features and deep learning[J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203009. (in Chinese)
- [4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, 1: 580-587.
- [5] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, 39(6): 1137-1149.
- [6] Redmon J, Farhadi A. YOLOv3: An incremental improvement [EB/OL].[2019-04-01]. <https://arxiv.org/abs/1804.02767>.
- [7] Terrail J O D, Jurie, Frédéric. Faster RER-CNN: application to the detection of vehicles in aerial images [EB/OL].[2019-04-01]. <https://arxiv.org/abs/1809.07628>.
- [8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 1:770-778.
- [9] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 1: 4510-4520.
- [10] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices [C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2018, 1: 6848-6856.